

****Titre du Cours : Introduction à Spark et Scala pour la Data Science****

****Objectifs du Cours :****

- Comprendre les concepts fondamentaux de Spark et Scala.
- Savoir comment utiliser Spark pour le traitement de données à grande échelle.
- Acquérir des compétences pratiques pour l'analyse de données en utilisant Spark et Scala.

Les points à aborder (essayez d'aborder le maximum de points) :

1. Introduction à Spark et Scala

- Présentation du cours et des objectifs.
- Introduction à Spark : Qu'est-ce que Spark, avantages et cas d'utilisation.
- Introduction à Scala : Les bases de la programmation en Scala.
- Installation de Spark et Scala sur un environnement local.

2. RDD (Resilient Distributed Datasets)

- Comprendre les RDD, l'unité fondamentale de Spark.
- Création de RDD, transformations et actions.
- Exercices pratiques avec RDD.

3. Spark SQL et DataFrames

- Introduction à Spark SQL.
- Utilisation de DataFrames pour le traitement de données structurées.
- Requêtes SQL avec Spark.
- Exercices pratiques avec DataFrames.

4. Spark Streaming

- Introduction à Spark Streaming pour le traitement de données en continu.
- Traitement de données en temps réel avec Spark Streaming.
- Exercices pratiques avec Spark Streaming.

5. Machine Learning avec Spark MLlib

- Présentation de MLlib, la bibliothèque de machine learning de Spark.
- Apprentissage supervisé et non supervisé avec MLlib.
- Exercices pratiques de machine learning avec Spark.

6. GraphX

- Introduction à GraphX pour le traitement de données en graphes.
- Manipulation de graphes avec GraphX.
- Exercices pratiques avec GraphX

7. Stockage et gestion de données

- Utilisation de Spark pour la lecture/écriture de données depuis/vers diverses sources de données (HDFS, Cassandra, etc.).
- Gestion de données massives avec Spark.
- Exercices pratiques sur la gestion de données.

8. Optimisation de performances

- Techniques d'optimisation pour les jobs Spark.
- Stratégies de tuning et de cache.
- Exercices pratiques d'optimisation.